

Pathways to Desistance

Datasheet

I. MOTIVATION

I-A For what purpose was the dataset created?

“The larger goals of the Pathways study are to improve decision-making by court and social service personnel and to clarify policy debates about alternatives for serious adolescent offenders. We hope to provide juvenile justice professionals and policy-makers with reliable empirical information that can be applied to improve practice, particularly regarding juveniles’ competence and culpability, risk for future offending, and amenability to rehabilitation” [1].

I-B Who created the dataset?

Is it an official law enforcement or government body? An academic research team? Other?

The Pathways to Desistance study grew out of the planning efforts of the MacArthur Foundation Research Network on Adolescent Development and Juvenile Justice. Network activities provided the initial forum for conceptualizing and planning this study. Additional funding from an array of both federal and private agencies supported data collection and other study activities [1]. A full list of contributors can be found here:

<https://www.pathwaysstudy.pitt.edu/people.html>

I-C Was there a specific task in mind, or gap that needed to be filled?

“The aims of the investigation are to: identify initial patterns of how serious adolescent offenders stop antisocial activity; describe the role of social context and developmental changes in promoting these positive changes; and compare the effects of sanctions and interventions in promoting these changes” [1].

“Some commentators have questioned whether a separate juvenile justice system is even warranted, given its dismal record at controlling or deterring juvenile crime. This debate is occurring, however, with limited data on either patterns of desistance or escalation among serious adolescent offenders or the effects of interventions and sanctions on trajectories of offending during and after adolescence. Although some studies suggest that most offenders curtail or stop antisocial behavior in late adolescence, this research has relied on very small samples of serious offenders or on very limited measurement of antisocial behavior patterns and developmental change’ [1].

II. COMPOSITION

II-A What do the instances that comprise the dataset represent?

For example: crimes, offenders, court cases, police officers

Interview responses of youth offenders. Each participant was interviewed multiple times. Each interview is a different data instance.

II-B Are there multiple types of instances?

For example: offenders, victims, and the relationship between them.

Yes. In addition to interview responses, there are official records, e.g. of arrests, and other collateral information to verify the self-reported information.

II-C How many instances are there in total?

Of each type, if appropriate.

The dataset contains information on 1354 serious juvenile offenders. Each participant was followed for a period of seven years, with interviews conducted every 6 months for the first 3 years and every 12 months thereafter.

II-D Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?

Enrollment into the Pathways to Desistance study occurred over a twenty-six month period between November, 2000 and January, 2003.

To be eligible for the study, individuals had to be in Maricopa County, AZ or Philadelphia, PA and:

1. at least 14 years old and under 18 years old at the time of their committing offense.
2. found guilty of a serious offense (predominantly felonies, with a few exceptions for some misdemeanor property offenses, sexual assault, or weapons offenses).
3. had to provide informed assent or consent (parent consent was obtained for all youth under the age of 18 at the time of enrollment).

The proportion of male youth found guilty of a drug charge was capped at 15% to avoid an over-representation of drug offenders. All females who met the age and crime criteria were approached for enrollment as were youth being considered for trial in the adult system. Twenty percent of the youths approached for participation declined [1].

II-E What data does each instance consist of?

If there is a large number of variables, please provide a broad description of what is included.

Interview responses. In addition, official arrest and court records were obtained for each participant. Among other topics, participants were asked about their offending, interactions with the justice system, and alcohol and drug use.

Relevant to criminal justice, participants self-report their levels of offending for various categories. Specifically, participants are asked about the frequency of committing each of the following acts over the past year (first interview) or from the last interview: Destroy property, set fire, broke in to steal, shoplift, receive stolen prop, use credit card illegally, stole car, sold marijuana, sold other drug, carjacked, drove drunk, been paid by someone for sex, forced sex, killed someone, shot someone, shot at someone, robbery with weapon, robbery no weapon, beaten someone, in fight, fight part of gang, carried gun, enter car to steal, gone joyriding.

Data of re-arrests from official records is also reported.

For full details please see:

<https://www.pathwaysstudy.pitt.edu/codebook/sro-sb.html>.

II-F Is there a target label or associated with each instance?

Please include labels that are likely to be used as target labels, e.g. recidivism.

No. However, re-offending or re-arrest may be suitable to be used as target labels.

II-G Are there recommended data splits (e.g., training, development/validation, testing)?

If so, please provide a description of these splits, explaining the rationale behind them.

No.

II-H Does the dataset contain data on race and ethnicity?

If so, is it based on the individual's self-description, or based on officer's impression? Was it collected or derived in post-processing? For example, by name analysis.

Yes. This information is self-reported by the participants.

II-I Are there any known errors, sources of noise, bias or missing data, or variables collected for only part of the datasets?

If so, please provide a description.

The data is self-reported. Although efforts were made to corroborate and validate the information through various means, including interviews with others who know the participants and comparison to official arrest and court records.

The participants in this study are not a representative sample of the general population, and any findings might not be generalizable.

II-J Does the dataset contain data on criminal history or other data that might be considered confidential or sensitive in any way?

For example: sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; If so, please provide a description.

Yes. The survey contains information on criminal activity, alcohol and drug use/abuse, health including mental, domestic violence, relationships, psychological traits and IQ, opinions, religion, income, and demographic information.

II-K Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset?

If so, please describe how.

III. USES

III-A What type of tasks, if any, has the dataset been used for?

If so, please provide examples and include citations.

The findings of the original study can be found in [2], [3], [4].

III-B Is there a repository that links to any or all papers or systems that use the dataset?

If so, please provide a link or other access point.

Yes. Please see:

<https://www.pathwaysstudy.pitt.edu/publications.html>

III-C What (other) tasks could the dataset be used for?

This dataset can be used to investigate the relationship between offending and arrests, including conditioning on several demographic factors.

III-D Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?

Limitations include the small size of the sample and that it is non-representative of the general population.

IV. COLLECTION PROCESS

IV-A How was the data associated with each instance acquired?

e.g. the data collected survey, the raw data is routinely collected by the courts.

Interviews were done with participants. Collateral interviews were conducted with family members or peers. Official records were gathered regarding arrest and social service involvement [2].

IV-B *Was the information self-reported?*

If the data was self-reported, was the data validated/verified?
If so, please describe how.

Yes. But the information was corroborated via interviews with family members or peers and via official records whenever possible.

IV-C *Who was involved in the data collection process?*

Was this done as part of their other duties? If not, were they compensated?

Participants, who are serious juvenile offenders, and their family members and peers. Participants were paid between \$50 and \$150 for each interview [2].

IV-D *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)?*

If not, please describe the timeframe in which the data associated with the instances was created. If the collection was not continuous within the timeframe, please specify the intervals, for example, annually, every 4 years, irregularly.

Participants were recruited between the years 2000 – 2003. Each participant was followed for a period of 7 years.

IV-E *Were any ethical review processes conducted (e.g., by an institutional review board)?*

If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

Unknown.

IV-F *Were the individuals in question notified about the data collection? Did they give their consent?*

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?

Yes. Participation in the study was voluntary.

IV-G *Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted?*

If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

Unknown.

V. PRE-PROCESSING, CLEANING, LABELING

V-A *Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, removal of instances, processing of missing values)?*

If so, please provide a description and reference to the documentation. If not, you may skip the remaining questions in this section.

The technical report [2] does not mention data processing.

VI. DISTRIBUTION

VI-A *Is the data publicly available? How and where can it be accessed (e.g., website, GitHub)?*

Does the dataset have a digital object identifier (DOI)?

A version of the data, with some variables restricted is publicly available and can be accessed from here:

<https://www.icpsr.umich.edu/web/ICPSR/studies/36800>

VI-B *Is the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?*

If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

The data is in the public domain. Some variables are restricted are required requesting access.

The license is not specified, but a citation and deposit requirement are listed:

Citation Requirement: Publications based on ICPSR data collections should acknowledge those sources by means of bibliographic citations. To ensure that such source attributions are captured for social science bibliographic utilities, citations must appear in footnotes or in the reference section of publications.

Deposit Requirement: To provide funding agencies with essential information about use of archival resources and to facilitate the exchange of information about ICPSR participants' research activities, users of ICPSR data are requested to send to ICPSR bibliographic citations for each completed manuscript or thesis abstract. Visit the ICPSR Web site for more information on submitting citations.

VII. MAINTENANCE

VII-A *Is the dataset maintained? Who is supporting/hosting/maintaining the dataset?*

The dataset has a website that is maintained by the Center for Research on Health Care (CRHC) Data Center.

VII-B *How can the owner/curator/manager of the dataset be contacted (e.g., email address)?*

Please see website for up to data contact information:

<https://www.pathwaysstudy.pitt.edu/contactPage/contact.aspx>

VII-C *Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?*

No.

VII-D *Are older versions of the dataset continue to be supported/hosted/maintained?*

N/A

VII-E *If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?*

If so, please provide a description.

No.

REFERENCES

- [1] "Pathways to Desistance."
- [2] "Pathways to Desistance – Final Technical Report."
- [3] E. Mulvey, L. Steinberg, J. Fagan, E. Cauffman, A. Piquero, L. Chassin, G. Knight, R. Brame, C. Schubert, T. Hecker, and S. Losoya, "Theory and research on desistance from antisocial activity among serious adolescent offenders," *Youth Violence and Juvenile Justice*, vol. 2, no. 3, pp. 213–236, Jul. 2004.
- [4] E. P. Mulvey and C. A. Schubert, "Some Initial Findings and Policy Implications of the Pathways to Desistance Study," *Victims & Offenders*, vol. 7, no. 4, pp. 407–427, 2012.